

Precision Tracing of Household Dengue Spread Using Inter- and Intra-Host Viral Variation Data, Kamphaeng Phet, Thailand

Irina Maljkovic Berry,¹ Melanie C. Melendrez,¹ Simon Pollett, Katherine Figueroa, Darunee Buddhari, Chonticha Klunghthong, Ananda Nisalak,² Michael Panciera, Butsaya Thaisomboonsuk, Tao Li, Tyghe G. Vallard, Louis Macareo, In-Kyu Yoon, Stephen J. Thomas, Timothy Endy, Richard G. Jarman

Dengue control approaches are best informed by granular spatial epidemiology of these viruses, yet reconstruction of inter- and intra-household transmissions is limited when analyzing case count, serologic, or genomic consensus sequence data. To determine viral spread on a finer spatial scale, we extended phylogenomic discrete trait analyses to reconstructions of house-to-house transmissions within a prospective cluster study in Kamphaeng Phet, Thailand. For additional resolution and transmission confirmation, we mapped dengue intra-host single nucleotide variants on the taxa of these time-scaled phylogenies. This approach confirmed 19 household transmissions and revealed that dengue disperses an average of 70 m per day between households in these communities. We describe an evolutionary biology framework for the resolution of dengue transmissions that cannot be differentiated based on epidemiologic and consensus genome data alone. This framework can be used as a public health tool to inform control approaches and enable precise tracing of dengue transmissions.

Dengue virus (DENV) causes an estimated 390 million infections each year, 96 million of which manifest as clinical disease (1). Dengue is endemic in >100 countries. An estimated 3.9 billion persons are at risk for infection, of which 75% reside in the Asia-Pacific region (2). DENV control approaches are best informed by granular spatial epidemiology of these

viruses, and investigators have long tried to understand the landscape of DENV dynamics and spread. Robust public health surveillance and rigorous academic research have enabled tracking of clinical infections to understand changing demographics, populations at risk, hyperendemicity, transmission, disease severity, and virus dispersal patterns within and across human populations (3–10). With the advent of sequencing technologies, single-gene and whole-genome DENV analyses have complemented case-based and serologic surveillance and enabled further insights into DENV epidemic dynamics and spread, disease severity, introductions and emergence of novel variants, tracking of DENV transmissions, and monitoring of viral diversity and evolution (11–17). Investigations of DENV infection dynamics, from the scales of countries and districts down to the city, village, school, and house level, have been performed to describe the patterns and predictors of DENV spread (7,9,11–13,18–21). These studies have enabled many novel insights into DENV transmission and contributed to improved prediction, prevention, and control strategies. Such studies have also emphasized a key limitation of DENV genomic epidemiology: reconstructing transmission chains between households and within households is typically not possible, even though resolving DENV spread at such fine scales would offer major relevance to public health response.

Consensus whole-genome sequences, which are typically used in genomic studies of viral epidemics and spread, typically have insufficient variability to distinguish between infecting strains sampled within 2 weeks of each other, especially if closely sampled in space (22). This limitation might lead to unresolved

Author affiliations: Walter Reed Army Institute of Research Viral Diseases Branch, Silver Spring, Maryland, USA (I. Maljkovic Berry, M.C. Melendrez, S. Pollett, K. Figueroa, M. Panciera, T. Li, T.G. Vallard, R.G. Jarman); Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand (D. Buddhari, C. Klunghthong, A. Nisalak, B. Thaisomboonsuk, L. Macareo); International Vaccine Institute, Seoul, South Korea (I.-K. Yoon); Upstate Medical University of New York Department of Medicine, Syracuse, New York, USA (S.J. Thomas, T. Endy)

DOI: <https://doi.org/10.3201/eid2706.204323>

¹These authors contributed equally to this article.

²Deceased.

and low confidence phylogenetic tree topologies, unable to discern the exact relationships between closely related taxa, and thus unable to provide confident insights into the fine scale viral transmission patterns. In the case of DENV, the use of consensus env-gene or whole-genome data has been able to resolve fine-scale clustering of cases within 200 m but has not been able to determine discrete inter-household or within-household transmissions (12).

DENV, like other RNA viruses, exists within a host as a population of distinct viral variants. Together, these viral variants are usually assembled into a single viral consensus genome, but separately, they hold additional intra-host variant sequence information that can provide further resolution. Transmission of within-host minor viral variants, existing at low frequencies and therefore often not reflected in the consensus genome, has been reported for several different viral pathogens (23–27). An increasing number of studies have used this additional genetic information to confirm viral transmissions, and new tools that use both intra- and inter-host genetic variation have been developed for inference of transmission chains and transmission directions (23,26,28,29). We designed a study that explicitly measures the patterns of DENV minor variant transmission in a natural epidemic setting, to resolve more spatially explicit viral transmissions.

We investigated the level of transmission reconstruction granularity that can be achieved for DENV by analyzing 410 DENV whole genomes sampled from human and vector specimens in Kamphaeng Phet, Thailand, and sequenced by using next-generation sequencing techniques. A combination of Bayesian analyses and intra-host single nucleotide variant (iSNV) information, along with temporal and spatial epidemiologic data, enabled fine-scale reconstructions of transmission chains and house-to-house spread, as well as genomic confirmation of within-household transmission clusters. This approach could be used in public health to reconstruct fine-scale DENV transmissions and support dengue control and prevention efforts.

Methods

Ethics Statement

The samples used in this study were virus isolates, which originated from samples collected in a study described in Thomas et al. (30), per the protocols approved by the institutional review boards (IRBs) of the Thai Ministry of Public Health, Walter Reed Army Institute of Research, and the State University

of New York's Upstate Medical University. The IRBs of the University of California, Davis, University of Rhode Island, and University at Buffalo established relying agreements with the Walter Reed Army Institute of Research IRB. All isolates were deidentified, and this study's research team had no access to identification codes.

Prospective Cluster Study and Data Collection

The prospective cluster study method has been described previously (30). In brief, the study was conducted in Kamphaeng Phet Province (Kamphaeng Phet), Thailand. A blood sample was obtained from patients admitted to Kamphaeng Phet Hospital with a diagnosis of an acute dengue infection; if positive for DENV by PCR, the virus was isolated with low passage number in C6/36 cell culture or in *Toxorhynchites splendens* mosquito followed by C6/36 (Appendix Table 1, <https://wwwnc.cdc.gov/EID/article/27/6/20-4323-App1.pdf>). Patients who were DENV PCR-positive were considered index case-patients for a cluster investigation. The exact house locations of index case-patients was mapped through a geographic information system by using a global positioning system unit. Homes within 100–200 m radius around the index case-patient's house were screened for additional cases in persons with a history of temperature $\geq 38^{\circ}\text{C}$ within past 7 days, and a blood sample was collected from those persons who agreed to participate in the study. DENV reverse transcription PCR, viral isolation, and IgM/IgG ELISA were performed on all collected specimens.

Sequencing, Genome Assembly, and Minor Variant Determination

We sequenced all DENV isolates to obtain whole genomes during 2009–2012 by using the Roche 454 FLX system (Roche, <https://www.roche.com>), the Illumina MiSeq next-generation sequencing system (Illumina, <https://www.illumina.com>), and, for gap filling, the Applied Biosystems 3130 Sanger sequencing platform (ThermoFisher Scientific, <https://www.thermofisher.com>). We generated consensus genomes by using the in-house developed ngs_mapper reference mapping pipeline and manually curated to ensure consensus accuracy (31). We submitted all consensus genomes to GenBank (accession nos. MN448597–9006). In addition to consensus, we screened all DENV serotype 1 (DENV-1) MiSeq-derived assemblies that had a minimum 1,000× depth of coverage throughout the genome for presence of iSNVs. We excluded DENV serotype 2 (DENV-2) genomes from iSNV analyses because some of the

samples did not have enough coverage, mainly because of being sequenced by using Roche 454. iSNVs were called if they were present at a conservative frequency of $\geq 1\%$ (minimum Phred of 30) called by Lofreq, a base caller previously shown to confidently call DENV iSNVs at this frequency (32,33). Additional manual iSNV curation removed variants present because of primer induced error, certain types of sequencing error, and strand bias higher than what was observed in confident iSNVs from each genome. In addition, we manually removed iSNVs that were consistently present at either ends of the reads from the analyses because these iSNVs have previously been shown to be spurious (33).

Phylogenetic Analyses

We combined genomes for DENV serotypes 1–4 with GenBank references and aligned by using MUSCLE 3.8 (34). We used jModeltest2 to determine the best fit model of nucleotide substitution (35) and constructed maximum-likelihood trees for whole-genome sequence data from all 4 serotypes by using PhyML 3.0 with aLRT node support (36). Because discrete phylogeographic analyses to infer the household-scale geographic histories of all sampled viruses were computationally unfeasible, we sub-selected datasets from DENV maximum-likelihood trees that had most taxa sampled from households in the same subdistrict and had sufficient temporal structure (root-tip regression $r > 0.7$) as determined by Temp-Est (37). Using these criteria, we selected 2 DENV-1 sublineages and 3 DENV-2 sublineages to infer the inter-household patterns of DENV spread by using BEAST 1.8.4 (38). We excluded DENV serotypes 3 and 4 (DENV-3 and DENV-4) from further analyses because of the weak sublineage temporal structure and limited clade sizes by individual subdistricts of Kamphaeng Phet. For each sublineage analysis, we geotagged time-stamped taxa by household as discrete traits. The final Markov Chain Monte Carlo chains had lengths required for statistical convergence, as indicated by effective sample size values > 200 for key evolutionary parameters (Appendix Table 2). We constructed and annotated maximum clade credibility (MCC) trees by using Tree-Annotator. We manually inspected the MCC trees to determine probable and possible inter-household transmissions events. We defined probable inter-household transmissions as origin household being directly ancestral to the destination household and both geographic states supported with a probability ≥ 0.8 and plotted all iSNVs onto the resulting MCC trees. We used this information for confirmation or

detection of between-household spread and for detection of within-household connections.

iSNV Distribution Statistical Analyses

We compared the frequency of infected case-patients sharing iSNVs within a transmission cluster with the frequency of samples sharing these iSNVs across the sublineage by using χ^2 test. We defined a transmission cluster as the smallest cluster of genomes in the MCC tree containing all the genomes sharing the same iSNVs. Thus, if shared iSNVs are distributed across the tree in an unrelated manner, the transmission cluster they define will be large and will not differ significantly from the whole sublineage tree. A significant difference from the sublineage tree would thus indicate iSNV clustering not expected by chance. In addition, we compared pairwise p-distances between the genomes sharing iSNVs with 100 replicates of pairwise p-distances between 3 randomly sampled genomes across the 2 DENV-1 sublineages by using an in-house developed script. Using 3 genome pairwise distance in the randomized dataset was determined on the basis of the average size of the significant transmission clusters, as defined by χ^2 test.

Linear Regression Analyses of Transmission-Pair Distance over Sampling Time

We performed crude estimates of the dispersal rate of DENV at an inter-household scale by a univariate linear regression model, which fit the difference in sampling days between confirmed transmission pairs as a single predictor, and distance between confirmed transmission pairs as the outcome variable. We fit this model by using all transmission pairs confirmed by the Bayesian consensus analysis, the iSNV analyses, or both. We used the coefficient of determination (R^2) to determine model fit and the coefficient point estimate and 95% CI to estimate the average rate of DENV spread between households. We performed all regression analyses by using Stata 15.1 (StataCorp, <https://www.stata.com>).

Results

PCR-confirmed DENV cases in Kamphaeng Phet during 2010–2012 were dominated by DENV-2, with co-circulating DENV-1 and DENV-3. DENV-4 was detected to a small degree in 2011 and 2012 (Appendix Figure 1). A total of 410 DENV whole genomes were sequenced from samples collected during 2009–2012 in Kamphaeng Phet, including 100 DENV-1, 233 DENV-2, 64 DENV-3, and 13 DENV-4 genomes. The unique design of the Kamphaeng Phet cluster study provided an opportunity to investigate fine-scale

DENV spread between households in Kamphaeng Phet. Analyses of DENV serotype 1–4 phylogenetic trees identified 2 DENV-1 and 3 DENV-2 sublineages with adequate temporal structure and sufficient sample size for Bayesian analyses. DENV consensus genomes from DENV-1 and DENV-2 sublineages geotagged by their respective household of sampling were used as discrete traits for Bayesian ancestral state reconstructions. We confirmed between-household direct connections with high probability in 4 DENV-1 cases and 3 DENV-2 cases, in 7 different subdistricts of Kamphaeng Phet (Table 1; Appendix Figures 2, 3).

These analyses also indicated many weakly supported inter-household transmissions (probability of origin or destination household <0.8) (Appendix Table 3). This finding highlighted the limited spatial and temporal resolution that is obtained by consensus genome data only, and prompted an analysis of iSNVs.

We determined iSNVs for each of the DENV-1 genomes and plotted on the sublineage MCC trees of DENV-1. Closely related viruses in the MCC phylogenies often shared the same iSNV spectra, including ≥2 shared iSNVs (Appendix, Appendix Figure 2). In total, this analysis revealed 3 transmission clusters that involved shared iSNVs, revealing direct DENV spread between households that were 20–800 m apart and involved a total of 11 persons (Table 2; Figure). The clusters contained persons from both separate and same households. The transmission confirmed by BEAST in transmission cluster 3 consisted of viruses with very little within-host variation (0 and 1 iSNV) (Appendix Figure 2), indicating that close transmission connections were not always characterized by iSNV sharing. In addition, this transmission cluster showed how iSNVs can be lost in bigger clusters over time, such as the loss of 1853-C/T iSNV in 1 person from household TN18H023 (Appendix Figure 2). Thus, the presence of ≥2 shared iSNVs indicated fine-scale virus transmission connections, whereas absence of ≥2 shared iSNVs did not imply lack of shared connection.

We examined the distribution of sampling times within the 3 iSNV-confirmed transmission clusters to further characterize DENV transmission dynamics at this fine scale. All persons from transmission cluster 3 were sampled within the same day, suggesting that they were inoculated at about the same time (Table 2). The close household proximity and time of infection, together with identical virus consensus sequence and shared minor variants, would indicate infection from a common source and simultaneous minor variant spread to ≥6 different persons. In contrast, persons from transmission cluster 1 were sampled ≈2 weeks apart, a period that exceeds the known DENV incubation period and would have required transmission and preservation of minor variants through several bottlenecks across the invertebrate–vertebrate transmission cycle. By comparison, household transmissions confirmed by BEAST analyses only (no shared iSNVs) were on average 19 days apart (Table 1).

Finally, we used the transmission events confirmed by either Bayesian consensus sequence analysis (Table 1) or iSNV analyses (Table 2) to estimate the rate of DENV spread within subdistricts and between households. We performed a linear regression of transmission-pair distances over sampling time, with a model fit indicating a strikingly linear relationship ($R^2 = 0.91$) and a regression coefficient indicating that DENV disperses an average of 70 m/day (95% CI 54–86 m/day) within subdistricts in this study population (Appendix Figure 4).

Discussion

Dengue control approaches are best informed by granular characterization of the spatial epidemiology of these viruses. Even in active surveillance cohorts, however, resolution of transmissions on a household-scale is hampered by the limited geographic resolution permitted by case count, serologic, or even DENV whole-genome consensus sequence data. Our study

Table 1. Summary of probable household-to-household spread of dengue virus inferred by consensus genomes, Kamphaeng Phet, Thailand, 2009–2012

Serotype and sublineage	Subdistrict	Origin household† (location probability)	Destination household (location probability)	Approximate distance, m	Sampling time difference, d
DENV-1, sublineage 1	LD	LD02H075 (1.0)	LD01H116 (1.0)	1,400	19
DENV-1, sublineage 1	SK	SK06H346 (0.99)	SK06H370 (1.0)	3,000	29
DENV-1, sublineage 7	TN	TN18H021 (0.88)	TN18H014 (1.0)	80	1
DENV-1, sublineage 7	NC	NC06H057 (0.82)	NC06H407 (1.0)	30	0, 14‡
DENV-2, sublineage 2	NB	NB06H055 (0.80)	NB06H084 (1.0)	800	13
DENV-2, sublineage 2	NP	NP08H080 (0.82)	NP08H044 (1.0)	180	0
DENV-2, sublineage 6	SK	SK06H790 (0.99)	SK06H485 (1.0)	3,000	49

*LD, Lan Dokmai; NB, Na Bo Kham; NC, Nakhon Chum; NP, Nong Pling; SK, Sa Kaeo; TN, Thap Nakhon.

†Households are described by their subdistrict, cluster, and house numbers, such that the first 2 letters denote subdistrict, the next 2-digit number denotes cluster, and an H followed by a 3-digit number denotes house number.

‡Two persons from the donor household were dengue virus–positive 2 weeks apart.‡

Table 2. Dengue virus serotype 1 transmission clusters revealed by minor variant sharing, Kamphaeng Phet, Thailand, 2009–2012*

Sublineage	Transmission cluster	Household†	No. persons	Sampling dates	Approximate distance, m
1	1	LD02H056	1	2011 Jun 2	800
		LD10H001	1	2011 Jun 16	
7	2	ST05H002	2	2011 Nov 28–29	0
7	3	TN18H023	3	2012 Oct 10	0–80
		TN18H019	1	2012 Oct 10	
		TN18H021‡	2	2012 Oct 10	
		TN18H014‡	1	2012 Oct 9	

*LD, Lan Dokmai; ST, Song Tham; TN, Thap Nakhon.

†Households are described by their subdistrict, cluster, and house numbers, such that the first 2 letters denote subdistrict, the next 2-digit number denotes cluster, and an H followed by a 3-digit number denotes house number.

‡House or individual involvement with the transmission cluster also confirmed by BEAST analyses (38).

leveraged a prospective active surveillance study with dense case sampling over 4 years, 410 whole-genome sequences and a unique combination of within-host and between-host viral genome variability to reconstruct transmission of DENV at unprecedented spatial scales. By using multiple lines of evidence, we tracked the transmission of DENV minor variants to trace DENV transmissions between households. Our findings emphasize the value of deep sequencing in resolving house-to-house transmission patterns, and

offer a proof-of-concept approach to use intra-host DENV diversity for fine-scale case linkage as a public health tool. Our approach also resolved 3 transmission chains involving persons residing in the same households, providing direct genomic evidence for peridomestic transmission of DENV. We therefore present an evolutionary biology framework for the resolution of DENV transmissions that cannot be differentiated on the basis of epidemiologic and consensus genome data alone.

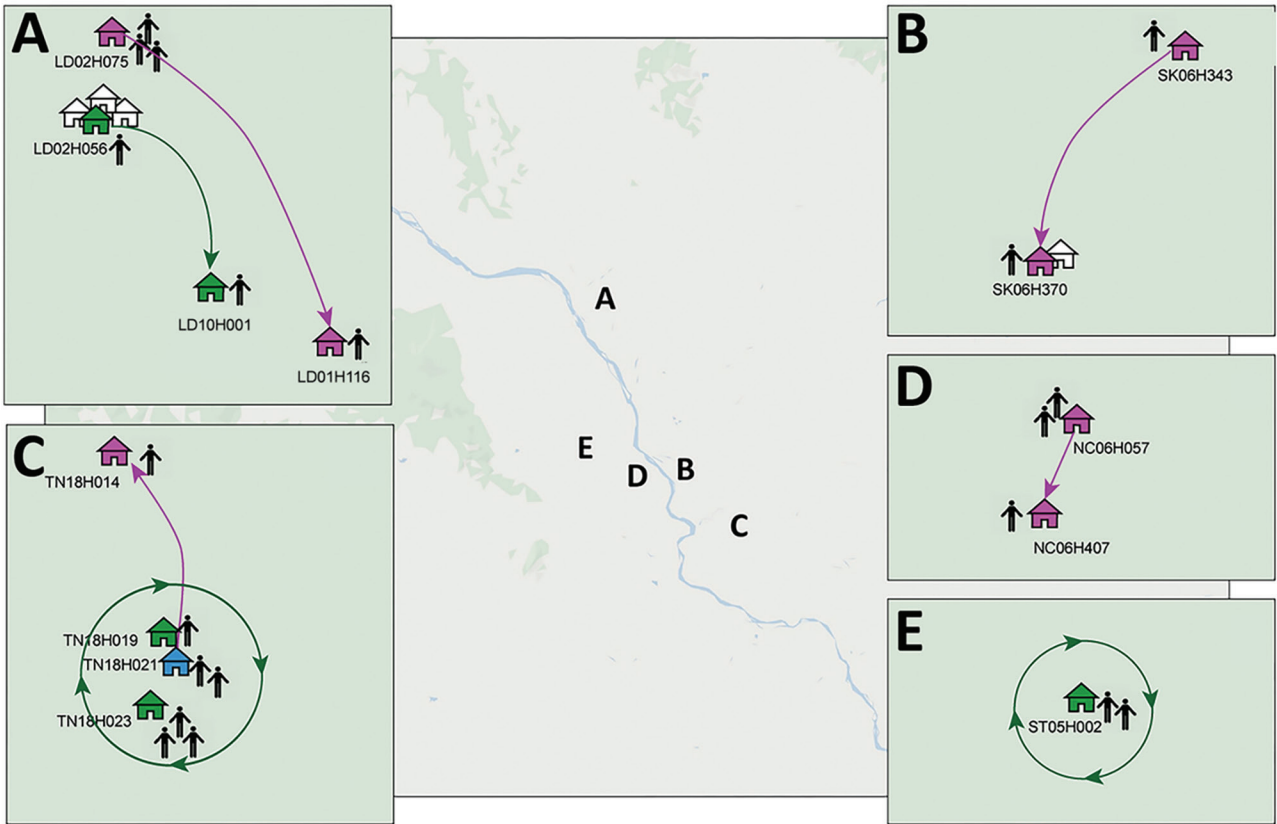


Figure. Approximate locations of dengue virus serotype 1 household transmission clusters and chains, Kamphaeng Phet, Thailand. Chains were confirmed by Bayesian consensus sequence (magenta houses) or minor variant (green houses) analyses or both methods (blue house). Households are described by their subdistrict, cluster, and house numbers, such that the first 2 letters denote subdistrict, the next 2-digit number denotes cluster, and an H followed by a 3-digit number denotes house number. LD, Lan Dokmai; NC, Nakhon Chum; SK, Sa Kaeo; ST, Song Tham; TN, Thap Nakhon.

Our results also show that sharing of iSNV patterns is not guaranteed between epidemiologically linked cases sampled within a 2 week period. iSNVs might be lost because of bottlenecks, genetic drift, or individual immune responses. Therefore, the transmission dynamics of iSNVs are multifactorial, and iSNV-based case-linkage should be performed carefully in the context of complementary epidemiologic and consensus genome data. This point is emphasized by our detection of single iSNVs shared among multiple unrelated cases across entire DENV lineages, suggesting that identical single iSNVs might be stochastically found among unlinked cases. Distinguishing these stochastic iSNVs from transmission-related iSNVs is essential, highlighting the importance of inclusion of closely related background datasets, as done in our study. Only a small proportion of iSNVs in our sequenced cases could be attributed to transmission events, similar to the findings of Sim et al (39), and this finding emphasizes the complex evolutionary landscapes of arbovirus iSNVs. Recently, methods have been developed to take into account intra- and inter-host variation for analyses of pathogen spread; however, the limited number of iSNVs in DENV, short sequencing reads, and erroneous or shared iSNVs across unrelated taxa, might be limiting factors influencing the confidence of these tools (40–42). Our findings also offer a framework to determine functional DENV bottleneck sizes between sampled human dengue cases. Prior studies have shown that bottlenecks within the mosquito vectors themselves range from 5 to 42 genomes and that as many as 11.5% iSNVs are shared between suspected human-human transmission pairs (39,43). Such bottleneck estimates broadly fit with our observations of constellations of as many as 5 iSNVs shared between human cases sampled 14 days apart. However, long-read sequencing and larger sample sizes, as well as direct samples, from both vector and human populations would be required to definitively determine the number of distinct viral genomes that are passed through human-vector bottlenecks. Even though our genomes were derived from low-passage isolates, we did not observe a significant correlation between the number of passages or passage history and shared iSNVs (Appendix Table 1). Many iSNVs are preserved during low passage of DENV-1 clinical samples; nevertheless, in such instances, care must be taken along with use of appropriate statistics to ensure their accuracy (44).

The analyses of our confirmed consensus and minor variant transmission clusters further revealed a

strong DENV spatial structure on an inter-household scale. Although household transmission pair distances varied from 0 to 3,000 m, our regression analysis of transmission pair distance over sampling time showed a strong linear relationship, indicating distance is a major determinant of DENV transmission at these ultra-fine spatial scales, and providing pilot estimates of the diffusion speed of DENV spread between households of 70 m/day (95% CI 54–86 m/day). Even though our regression analysis indicated a good model fit ($R^2 = 0.91$), these coefficient estimates are nevertheless preliminary and are subject to potential nonindependence of transmission pair distance measurements and would benefit from confirmation in greater sample sizes in other study populations. This distance dependent dispersal between households was not noted in a prior Singapore study, which plotted uncorrected consensus genetic distance against geographic distance and sampling times (21). This finding highlights the importance of phylogenetic correction for genomic-based DENV tracing. Taken together, these findings support prior research that suggested that DENV transmissions are highly localized on this microspatial scale and might inform vector-control operations (12).

Although our findings were derived from an intensive, 4-year active cluster investigation across a very well sampled geographic area, ascertainment bias was a potential weakness in our study. All sequences were either derived from index case-patients accessing care for febrile illness or their contacts who were only sampled in households within 100–200 m of the index case-patient. However, our dense sampling of index cases over a 4-year study period ensured a large number of sequenced specimens across the Kamphaeng Phet region across multiple cluster investigations. Two of the 9 genetically confirmed transmission pairs were detected between persons who were not sampled in the same index case-contact cluster, highlighting the incremental value of genomic data over epidemiologic data alone in DENV transmission tracing. Future improvements in deep sequencing methods (45), coupled with more intensive vector and host sampling, could greatly improve our understanding of the invertebrate-vertebrate DENV iSNV bottleneck size and better resolve DENV vector-human transmission dynamics.

In conclusion, we have provided proof-of-concept evidence to support a new evolutionary epidemiologic framework that incorporates both within-host and between-host viral variation to determine fine-scale DENV transmissions. Our framework offers precision tracing of DENV at inter- and intra-household

scales and leverages recent similar applications used for study of influenza and Ebola viruses (23,24). Our findings warrant larger studies in these and other dengue surveillance cohorts, as well as exploration of how this approach can be used for other arboviruses such as chikungunya and Zika viruses. Ideally, such studies should be coupled with advances in the accuracy of long-read deployable sequencing platforms that could permit near real-time intra-host variant construction (46).

Acknowledgments

We thank Henrik Salje and Alan Rothman for their advice.

This work was supported by the National Institute of Allergy and Infectious Diseases (grant no. R01 GM083224), the Military Infectious Diseases Program, and the Armed Forces Health Surveillance Branch and its Global Emerging Infections Surveillance Section.

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the National Institutes of Health, Department of the Army, Department of Defense, or the US Government. Several of the authors are US Government employees. This work was prepared as part of their official duties. Title 17 U.S.C. § 105 provides that “Copyright protection under this title is not available for any work of the United States Government.” Title 17 U.S.C. §101 defines a US Government work as a work prepared by a military service member or employee of the US Government as part of that person’s official duties. The investigators have adhered to the policies for protection of human subjects as prescribed in Army Regulations 70–25.

About the Author

Dr. Maljkovic Berry is the chief of the Viral Genomics section at the Viral Diseases Branch, Walter Reed Army Institute of Research. Her research includes evolution of vectorborne and respiratory viruses as well as viral genomic surveillance.

References

- Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. 2013;496:504–7. <https://doi.org/10.1038/nature12060>
- Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, et al. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl Trop Dis*. 2012;6:e1760. <https://doi.org/10.1371/journal.pntd.0001760>
- Yoon IK. Focal dengue virus transmission in Kamphaeng Phet, Thailand and implications for management. *Southeast Asian J Trop Med Public Health*. 2015;46(Suppl 1):17–25.
- Alera MT, Srikiatkachorn A, Velasco JM, Tac-An IA, Lago CB, Clapham HE, et al. Incidence of dengue virus infection in adults and children in a prospective longitudinal cohort in the Philippines. *PLoS Negl Trop Dis*. 2016;10:e0004337. <https://doi.org/10.1371/journal.pntd.0004337>
- Nisalak A, Clapham HE, Kalayanaroj S, Klungthong C, Thaisomboonsuk B, Fernandez S, et al. Forty years of dengue surveillance at a tertiary pediatric hospital in Bangkok, Thailand, 1973–2012. *Am J Trop Med Hyg*. 2016;94:1342–7. <https://doi.org/10.4269/ajtmh.15-0337>
- Cucunawangsih, Lugito NPH. Trends of Dengue Disease Epidemiology. *Virology (Auckl)*. 2017;8:1178122X17695836.
- Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, et al. Global spread of dengue virus types: mapping the 70 year history. *Trends Microbiol*. 2014;22:138–46. <https://doi.org/10.1016/j.tim.2013.12.011>
- Hoang Quoc C, Henrik S, Isabel RB, In-Kyu Y, Chau NV, Hung NT, et al. Synchrony of dengue incidence in Ho Chi Minh City and Bangkok. *PLoS Negl Trop Dis*. 2016;10:e0005188. <https://doi.org/10.1371/journal.pntd.0005188>
- Bhoomiboonchoo P, Gibbons RV, Huang A, Yoon IK, Buddhari D, Nisalak A, et al. The spatial dynamics of dengue virus in Kamphaeng Phet, Thailand. *PLoS Negl Trop Dis*. 2014;8:e3138. <https://doi.org/10.1371/journal.pntd.0003138>
- Bhoomiboonchoo P, Nisalak A, Chansatiporn N, Yoon IK, Kalayanaroj S, Thipayamongkolgul M, et al. Sequential dengue virus infections detected in active and passive surveillance programs in Thailand, 1994–2010. *BMC Public Health*. 2015;15:250. <https://doi.org/10.1186/s12889-015-1590-z>
- Pollett S, Melendrez MC, Maljkovic Berry I, Duchêne S, Salje H, Cummings DAT, et al. Understanding dengue virus evolution to support epidemic surveillance and counter-measure development. *Infect Genet Evol*. 2018;62:279–95. <https://doi.org/10.1016/j.meegid.2018.04.032>
- Salje H, Lessler J, Maljkovic Berry I, Melendrez MC, Endy T, Kalayanaroj S, et al. Dengue diversity across spatial and temporal scales: Local structure and the effect of host population size. *Science*. 2017;355:1302–6. <https://doi.org/10.1126/science.aaj9384>
- Tian H, Sun Z, Faria NR, Yang J, Cazelles B, Huang S, et al. Increasing airline travel may facilitate co-circulation of multiple dengue virus serotypes in Asia. *PLoS Negl Trop Dis*. 2017;11:e0005694. <https://doi.org/10.1371/journal.pntd.0005694>
- Normile D. Tropical medicine. Surprising new dengue virus throws a spanner in disease control efforts. *Science*. 2013;342:415. <https://doi.org/10.1126/science.342.6157.415>
- Vasilakis N, Cardosa J, Hanley KA, Holmes EC, Weaver SC. Fever from the forest: prospects for the continued emergence of sylvatic dengue virus and its impact on public health. *Nat Rev Microbiol*. 2011;9:532–41. <https://doi.org/10.1038/nrmicro2595>
- Myat Thu H, Lowry K, Jiang L, Hlaing T, Holmes EC, Aaskov J. Lineage extinction and replacement in dengue type 1 virus populations are due to stochastic events rather than to natural selection. *Virology*. 2005;336:163–72. <https://doi.org/10.1016/j.virol.2005.03.018>
- Rico-Hesse R, Harrison LM, Salas RA, Tovar D, Nisalak A, Ramos C, et al. Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. *Virology*. 1997;230:244–51. <https://doi.org/10.1006/viro.1997.8504>
- Rabaa MA, Klungthong C, Yoon IK, Holmes EC, Chinnawirotpisan P, Thaisomboonsuk B, et al. Frequent in-migration and highly focal transmission of dengue viruses

- among children in Kamphaeng Phet, Thailand. *PLoS Negl Trop Dis*. 2013;7:e1990. <https://doi.org/10.1371/journal.pntd.0001990>
19. Rabaa MA, Ty Hang VT, Wills B, Farrar J, Simmons CP, Holmes EC. Phylogeography of recently emerged DENV-2 in southern Viet Nam. *PLoS Negl Trop Dis*. 2010;4:e766. <https://doi.org/10.1371/journal.pntd.0000766>
 20. Jarman RG, Holmes EC, Rodpradit P, Klungthong C, Gibbons RV, Nisalak A, et al. Microevolution of Dengue viruses circulating among primary school children in Kamphaeng Phet, Thailand. *J Virol*. 2008;82:5494–500. <https://doi.org/10.1128/JVI.02728-07>
 21. Schreiber MJ, Holmes EC, Ong SH, Soh HS, Liu W, Tanner L, et al. Genomic epidemiology of a dengue virus epidemic in urban Singapore. *J Virol*. 2009;83:4163–73. <https://doi.org/10.1128/JVI.02445-08>
 22. Pybus OG, Tatem AJ, Lemey P. Virus evolution and transmission in an ever more connected world. *Proc Biol Sci*. 2015;282:20142878. <https://doi.org/10.1098/rspb.2014.2878>
 23. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*. 2014;345:1369–72. <https://doi.org/10.1126/science.1259657>
 24. Stack JC, Murcia PR, Grenfell BT, Wood JL, Holmes EC. Inferring the inter-host transmission of influenza A virus using patterns of intra-host genetic variation. *Proc Biol Sci*. 2013;280:20122173. <https://doi.org/10.1098/rspb.2012.2173>
 25. Zanini F, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, et al. Population genomics of inpatient HIV-1 evolution. *eLife*. 2015;4:4. <https://doi.org/10.7554/eLife.11282>
 26. Fischer GE, Schaefer MK, Labus BJ, Sands L, Rowley P, Azzam IA, et al. Hepatitis C virus infections from unsafe injection practices at an endoscopy clinic in Las Vegas, Nevada, 2007–2008. *Clin Infect Dis*. 2010;51:267–73. <https://doi.org/10.1086/653937>
 27. Worby CJ, Lipsitch M, Hanage WP. Shared genomic variants: identification of transmission routes using pathogen deep-sequence data. *Am J Epidemiol*. 2017;186:1209–16. <https://doi.org/10.1093/aje/kwx182>
 28. Wymant C, Hall M, Ratmann O, Bonsall D, Golubchik T, de Cesare M, et al. PHYLOSCANNER: inferring transmission from within- and between-host pathogen genetic diversity. *Mol Biol Evol*. 2017.
 29. Skums P, Zelikovsky A, Singh R, Gussler W, Dimitrova Z, Knyazev S, et al. QUENTIN: reconstruction of disease transmissions from viral quasispecies genomic data. *Bioinformatics*. 2018;34:163–70. <https://doi.org/10.1093/bioinformatics/btx402>
 30. Thomas SJ, Aldstadt J, Jarman RG, Buddhari D, Yoon IK, Richardson JH, et al. Improving dengue virus capture rates in humans and vectors in Kamphaeng Phet Province, Thailand, using an enhanced spatiotemporal surveillance strategy. *Am J Trop Med Hyg*. 2015;93:24–32. <https://doi.org/10.4269/ajtmh.14-0242>
 31. Vallard T, Melendrez M, Panciera M. ngs_mapper v1.4.2: a pipeline for viral and microbial genome construction. 2014 [cited 2016 Mar 25]. https://github.com/VDBWRAIR/ngs_mapper
 32. Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res*. 2012;40:11189–201. <https://doi.org/10.1093/nar/gks918>
 33. McCrone JT, Lauring AS. Measurements of intrahost viral diversity are extremely sensitive to systematic errors in variant calling. *J Virol*. 2016;90:6884–95. <https://doi.org/10.1128/JVI.00667-16>
 34. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5:113. <https://doi.org/10.1186/1471-2105-5-113>
 35. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012;9:772. <https://doi.org/10.1038/nmeth.2109>
 36. Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol*. 2009;537:113–37. https://doi.org/10.1007/978-1-59745-251-9_6
 37. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*. 2016;2:vew007. <https://doi.org/10.1093/ve/vew007>
 38. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007;7:214. <https://doi.org/10.1186/1471-2148-7-214>
 39. Sim S, Aw PP, Wilm A, Teoh G, Hue KD, Nguyen NM, et al. Tracking dengue virus intra-host genetic diversity during human-to-mosquito transmission. *PLoS Negl Trop Dis*. 2015;9:e0004052. <https://doi.org/10.1371/journal.pntd.0004052>
 40. De Maio N, Worby CJ, Wilson DJ, Stoesser N. Bayesian reconstruction of transmission within outbreaks using genomic variants. *PLOS Comput Biol*. 2018;14:e1006117. <https://doi.org/10.1371/journal.pcbi.1006117>
 41. Alamil M, Hughes J, Berthier K, Desbiez C, Thébaud G, Soubeyrand S. Inferring epidemiological links from deep sequencing data: a statistical learning approach for human, animal and plant diseases. *Philos Trans R Soc Lond B Biol Sci*. 2019;374:20180258. <https://doi.org/10.1098/rstb.2018.0258>
 42. Didelot X, Fraser C, Gardy J, Colijn C. Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. *Mol Biol Evol*. 2017;34:997–1007. <https://doi.org/10.1093/molbev/msw275>
 43. Lequime S, Fontaine A, Ar Gouilh M, Moltini-Conclois I, Lambrechts L. Genetic drift, purifying selection and vector genotype shape dengue virus intra-host genetic diversity in mosquitoes. *PLoS Genet*. 2016;12:e1006111. <https://doi.org/10.1371/journal.pgen.1006111>
 44. Fung CK, Li T, Pollett S, Alera MT, Yoon IK, Hang J, et al. Effect of low-passage number on dengue consensus genomes and intra-host variant frequencies. *J Gen Virol*. 2021;102. <https://doi.org/10.1099/jgv.0.001553>
 45. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol*. 2019;20:8. <https://doi.org/10.1186/s13059-018-1618-7>
 46. Karst SM, Ziels RM, Kirkegaard RH, Sørensen EA, McDonald D, Zhu Q, et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat Methods*. 2021;18:165–9. <https://doi.org/10.1038/s41592-020-01041-y>

Address for correspondence: Irina Maljkovic Berry, Walter Reed Army Institute of Research, 503 Robert Grant Ave, Silver Spring, MD 20910, USA; email: irina.maljkovicberry.ctr@mail.mil

Precision Tracing of Household Dengue Spread Using Inter- and Intra-Host Viral Variation Data, Kamphaeng Phet, Thailand

Appendix

Genomes sharing one iSNV only were positioned across the phylogeny, even between relatively unrelated viruses, with no significant difference between the frequencies of samples sharing a single iSNV within a transmission cluster versus across the sublineage ($p = 0.38$). However, there was a significant difference between the frequencies of samples sharing two or more iSNVs occurring within a transmission cluster versus across the sublineage ($p = 0.013$), indicating that the distribution of strains sharing ≥ 2 iSNVs within transmission clusters was not due to chance. In addition, the consensus sequence p-distance between the samples sharing two or more iSNVs was significantly lower compared to the p-distance between randomly sampled genomes from the DENV1 sublineages ($p < 0.001$), indicating that the iSNV-sharing viruses were genetically significantly closer to each other than would be expected by chance. These results support the notion that these patterns of shared viral minor variants correlated with direct virus transmission chains and clusters. There was no significant correlation between the number of passages or passage history and shared minor variants (F-test, $p > 0.05$) (Appendix Table 1).

Appendix Table 1. DENV1 sublineage 1 and 7 sample passaging history, with transmission clusters (TC) highlighted in blue, and samples with shared iSNVs in red text.

D1 SL1		
Sample	Cell-passage no.	Transmission cluster
KDH0176A	C6/36-2	
KDH0168A	C6/36-2	
KDH0149A	C6/36-2	
KDH0131A	C6/36-3	
KDC0631A	C6/36-2	
KDH0142A	C6/36-3	
KDH0180A	C6/36-2	
KDH0163A	C6/36-2	
KDC0734A	C6/36-2	
KDH0154A	C6/36-2	
KDH0134A	TS-1, C6/36-1	
KDC0588A	C6/36-3	
KDC0587A	C6/36-3	
KDC0495A	C6/36-3	
KDH0151A	C6/36-2	
KDH0196A	TS-1, C6/36-1	
KDC0492A	C6/36-3	

D1 SL1		
Sample	Cell-passage no.	Transmission cluster
KDH0112A	TS-1,C6/36-1	
KDC0493A	C6/36-3	
KDH0143A	C6/36-2	
KDH0169A	TS-1,C6/36-1	
KDC0599A	C6/36-2	
KDC0501A	no info	
KDC0504A	no info	
KDH0132A	C6/36-3	
KDH0136A	TS-1,C6/36-1	
KDC0584A	C6/36-3	TC1
KDH0135A	C6/36-3	
KDH0129A	C6/36-3	
KDH0137A	TS-1,C6/36-1	
KDC0573A	C6/36-3	
KDC0574A	TS-1,C6/36-1	
KDC0574A2	TS-1,C6/36-1	
KDC0575A	C6/36-3	
KDH0110A	TS-1,C6/36-1	
KDC0455A	TS-1,C6/36-1	
KDC0454A	TS-1,C6/36-1	
KDH0102A	C6/36-3	
KDC0456A	C6/36-3	
KDH0100A	C6/36-3	
D1 SL7		
Sample	Cell-passage no.	Transmission cluster
KDH0081A	C6/36-3	
KDH0167A	C6/36-2	
KDC0722A	C6/36-2	
KDC1153A	C6/36-3	TC3
KDC1152A	C6/36-3	
KDC1151A	C6/36-3	
KDC1150A	C6/36-3	
KDC1149A	TS-1,C6/36-1	
KDC1148A	TS-1,C6/36-1	
KDH0309A	C6/36-3	
KDH0074A	TS-1, C6/36-1	
KDH0068A	C6/36-3	
KDC0394A	C6/36-3	
KDH0082A	C6/36-3	
KDH0087A	TS-1,C6/36-1	
KDH0073A	C6/36-4	
KDH0202A	TS-1,C6/36-1	
KDH0199A	C6/36-2	
KDH0198A	C6/36-2	
KDH0201A	TS-1,C6/36-1	TC2
KDC0885A	C6/36-2	
KDH0151A	C6/36-2	
KDH0160A	C6/36-2	
KDC0718A	C6/36-2	
KDC0640A	TS-1,C6/36-1	
KDH0144A	TS-1,C6/36-1	
KDH0194A	C6/36-3	

Appendix Table 2. BEAST parameters used for time-scaled inter-household spread analyses.

Serotype	Sublineage [†]	Nucleotide substitution model [‡]	Molecular clock model	Demographic model	Discrete trait (household) reconstruction model	MCMC length
DENV1	1	TrNef (TN93)	UCLN	Constant	CTMC, asymmetric	600 million
DENV1	7	TrNef (TN93)	UCLN	Constant	CTMC, asymmetric	1 billion
DENV2	2	GTR	UCLN	Constant	CTMC, asymmetric	600 million
DENV2	6	GTR+G+I	UCLN	Constant	CTMC, asymmetric	600 million
DENV2	11	TrNef (TN93)	UCLN	Constant	CTMC, asymmetric	600 million

[†]As classified in Appendix Figures 2 and 3.

[‡]Estimated by JModelTest2.

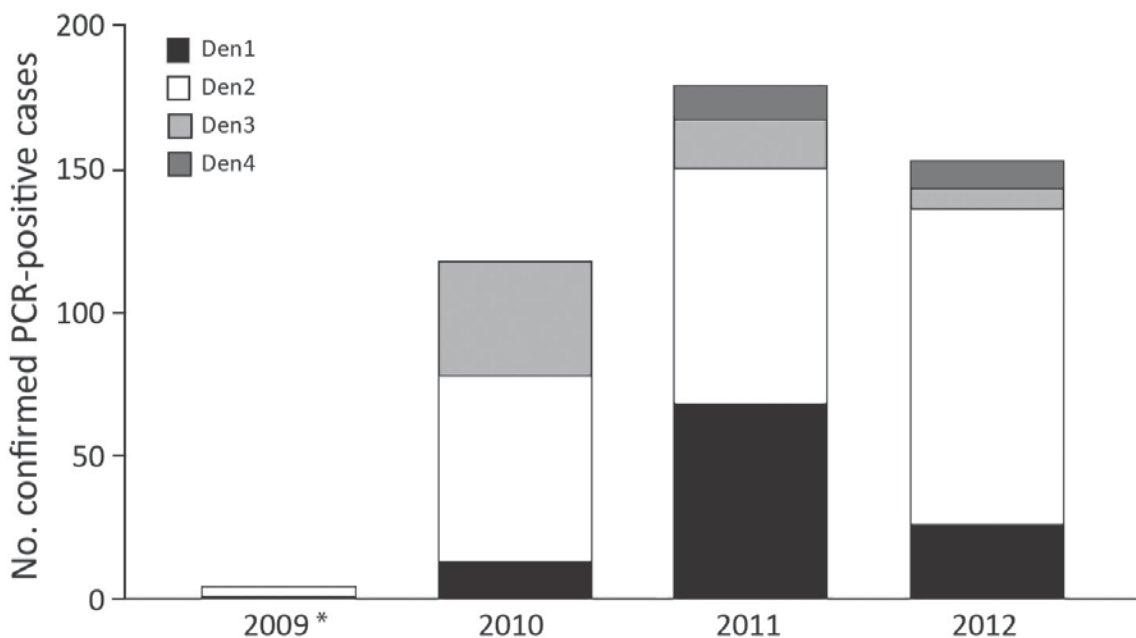
Appendix Table 3. Summary of probable and possible within-district household-to-household spread inferred by consensus genome discrete train Bayesian analysis†

Serotype	Sublineage	District	Origin household (probability)	Destination household (probability)	Level of evidence‡
D1	SL1	LD	LD02H002 (0.32)	LD02H033 (0.34)	Possible
D1	SL1	LD	LD02H033 (0.34)	LD02H001 (1.0)	Possible
D1	SL1	LD	LD02H002 (0.68)	LD02H008 (1.0)	Possible
D1	SL1	LD	LD02H002 (0.13)	LD01H001 (0.48)	Possible
D1	SL1	LD	LD01H001 (0.48)	LD02H075 (0.50)	Possible
D1	SL1	LD	LD02H075 (1.0)	LD01H116 (1.0)	Probable
D1	SL1	LD	LD02H075 (0.5)	LD02H056 (0.26)	Possible
D1	SL1	LD	LD02H056 (0.53)	LD02H085 (0.54)	Possible
D1	SL1	LD	LD02H085 (0.54)	LD10H001 (1.0)	Possible
D1	SL1	LD	LD02H084 (0.53)	LD02H021 (1.0)	Possible
D1	SL1	LD	LD02H056 (0.26)	LD02H084 (0.21)	Possible
D1	SL1	LD	LD02H084 (0.21)	LD10H044 (0.21)	Possible
D1	SL1	LD	LD02H084 (0.53)	LD10H021 (1.0)	Possible
D1	SL1	LD	LD02H044 (0.21)	LD10H042 (0.56)	Possible
D1	SL1	LD	LD10H042 (0.56)	LD02H098 (0.59)	Possible
D1	SL1	LD	LD10H044 (0.21)	LD10H043 (0.41)	Possible
D1	SL1	LD	LD10H043 (0.41)	LD10H044 (1.0)	Possible
D1	SL1	SK	SK10H028 (0.21)	SK12H001 (0.23)	Possible
D1	SL1	SK	SK12H001 (0.63)	SK11H001 (0.40)	Possible
D1	SL1	NM	NM24H098 (0.46)	NM24H107 (1.0)	Possible
D1	SL1	SK	SK12H001 (0.23)	SK10H081 (0.73)	Possible
D1	SL1	SK	SK10H081 (0.43)	SK06H346 (0.91)	Possible
D1	SL1	SK	SK06H346 (0.99)	SK06H370 (1.0)	Probable
D1	SL1	NC	NC09H028 (0.67)	NC09H029 (1.0)	Possible
D1	SL1	LD	LD01H073 (0.27)	LD01H051 (0.40)	Possible
D1	SL1	LD	LD01H051 (0.40)	LD01H037 (1.0)	Possible
D1	SL7	NC	NC12H016 (0.31)	NC12H001 (0.54)	Possible
D1	SL7	NC	NC12H016 (0.21)	NC06H057 (0.82)	Possible
D1	SL7	NC	NC12H001 (0.54)	NC04H179 (0.49)	Possible
D1	SL7	NC	NC06H057 (0.82)	NC06H407 (1.0)	Probable
D1	SL7	NC	NC04H179 (0.49)	NC10H001 (1.0)	Possible
D1	SL7	ST	ST05H002 (0.43)	ST04H001 (1.0)	Possible
D1	SL7	ST	ST05H002 (0.34)	ST12H072 (1.0)	Possible
D1	SL7	KT	KT03H049 (0.31)	KT03H371 (1.0)	Possible
D1	SL7	KT	KT03H054 (0.32)	KT03H049 (0.30)	Possible
D1	SL7	TN	TN18H021 (0.62)	TN18H023 (0.71)	Possible
D1	SL7	TN	TN18H023 (0.71)	TN18H019 (1.0)	Possible
D1	SL7	TN	TN18H021 (0.88)	TN18H014 (1.0)	Probable
D2	SL2	NB	NB06H055 (0.80)	NB06H084 (1.0)	Probable
D2	SL2	NP	NP08H080 (0.82)	NP08H044 (1.0)	Probable
D2	SL2	AT	AT12H032 (0.48)	AT12H026 (1.0)	Possible
D2	SL2	NP	NP07H043 (0.54)	NP07H051 (1.0)	Possible
D2	SL2	NB	NB09H001 (0.26)	NB09H026 (0.32)	Possible
D2	SL2	NB	NB09H026 (0.32)	NB09H004 (1.0)	Possible
D2	SL2	NB	NB09H026 (0.32)	NB09H008 (0.37)	Possible
D2	SL2	NB	NB09H008 (0.44)	NB09H026 (1.0)	Possible
D2	SL2	NB	NB09H008 (0.37)	NB21H030 (1.0)	Possible
D2	SL2	NP	NP07H319 (0.54)	NP08H001 (0.47)	Possible
D2	SL2	NP	NP08H001 (0.47)	NP08H080 (0.82)	Possible
D2	SL2	NB	NB10H001 (0.19)	NB06H046 (1.0)	Possible
D2	SL2	NB	NB10H001 (0.24)	NB18H015 (0.69)	Possible
D2	SL2	NB	NB18H015 (0.69)	NB18H091 (0.25)	Possible
D2	SL2	NB	NB18H091 (0.25)	NB10H040 (1.0)	Possible
D2	SL2	NB	NB18H001 (0.35)	NB06H073 (0.98)	Possible
D2	SL2	NB	NB18H001 (0.74)	NB01H028 (1.0)	Possible
D2	SL2	NB	NB10H001 (0.44)	NB10H079 (0.43)	Possible
D2	SL2	NB	NB18H079 (0.43)	NB18H054 (1.0)	Possible
D2	SL2	NB	NB18H079 (0.47)	NB10H001 (1.0)	Possible
D2	SL2	NB	NB18H001 (0.44)	NB18H089 (0.38)	Possible
D2	SL2	NB	NB18H089 (0.38)	NB06H056 (1.0)	Possible
D2	SL2	NB	NB18H089 (0.46)	NB06H055 (0.73)	Possible
D2	SL2	NB	NB06H055 (0.73)	NB10H068 (0.35)	Possible
D2	SL2	NB	NB06H068 (0.35)	NB10H099 (1.0)	Possible
D2	SL6	SK	SK06H790 (0.99)	SK06H485 (1.0)	Probable
D2	SL6	NB	NB18H030 (0.67)	NB04H086 (1.0)	Possible
D2	SL6	ST	ST06H010 (0.63)	ST02H030 (1.0)	Possible

Serotype	Sublineage	District	Origin household (probability)	Destination household (probability)	Level of evidence‡
D2	SL6	NM	NM25H016 (0.35)	NM03H095 (0.39)	Possible
D2	SL6	NM	NM03H095 (0.39)	NM19H055 (1.0)	Possible
D2	SL6	ST	ST06H010 (0.20)	ST04H032 (0.44)	Possible
D2	SL6	ST	ST04H032 (0.44)	ST04H043 (0.52)	Possible
D2	SL6	ST	ST04H043 (0.52)	ST07H013 (0.22)	Possible
D2	SL6	ST	ST07H013 (0.22)	ST07H003 (0.20)	Possible
D2	SL6	ST	ST07H003 (0.20)	STH7004 (0.72)	Possible
D2	SL6	ST	ST07H003 (0.31)	ST07H005 (1.0)	Possible
D2	SL6	ST	ST07H003 (0.57)	ST07H015 (1.0)	Possible
D2	SL6	ST	ST04H032 (0.57)	ST04H020 (0.79)	Possible
D2	SL6	ST	ST04H020 (0.79)	ST11H001 (0.48)	Possible
D2	SL6	ST	ST11H001 (0.48)	ST04H058 (1.0)	Possible
D2	SL6	ST	ST11H001 (0.77)	ST09H001 (0.62)	Possible
D2	SL6	ST	ST09H009 (0.36)	ST09H013 (1.0)	Possible
D2	SL11	TN	H05011055 (0.48)	H05111055 (1.0)	Possible
D2	SL11	NB	H01010063 (0.49)	H00610063 (1.0)	Possible
D2	SL11	NB	H00110063 (0.41)	H01010063 (0.49)	Possible

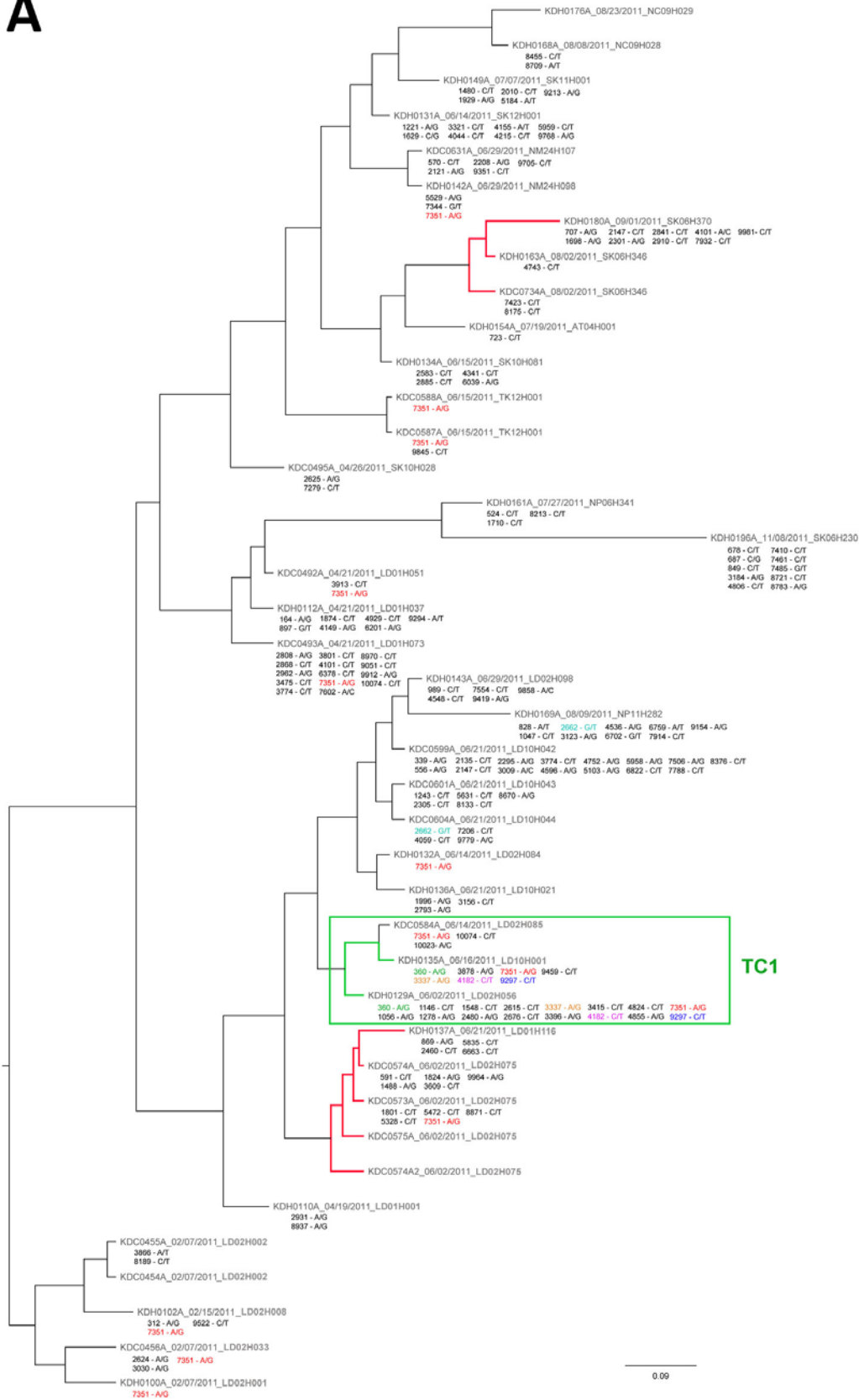
†Sublineages selected for analysis based on sufficient within-district sampling and adequate temporal structure (Appendix Figure 4)

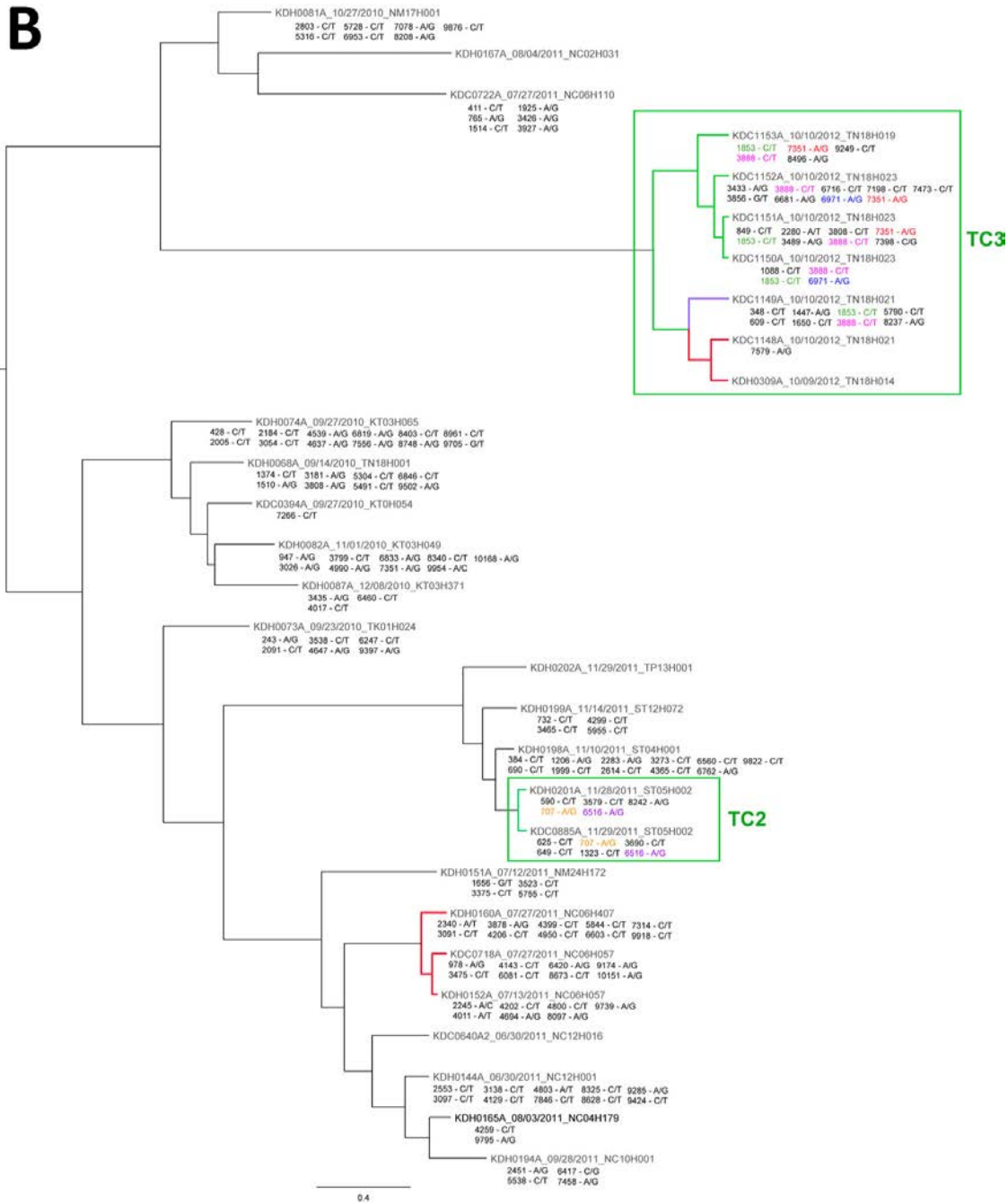
‡Probable defined as origin household being directly ancestral to the destination household and both geographic states supported with a probability ≥ 0.8 . Possible defined as origin household being directly ancestral to the destination household but with one or both geographic states having a probability < 0.8 .



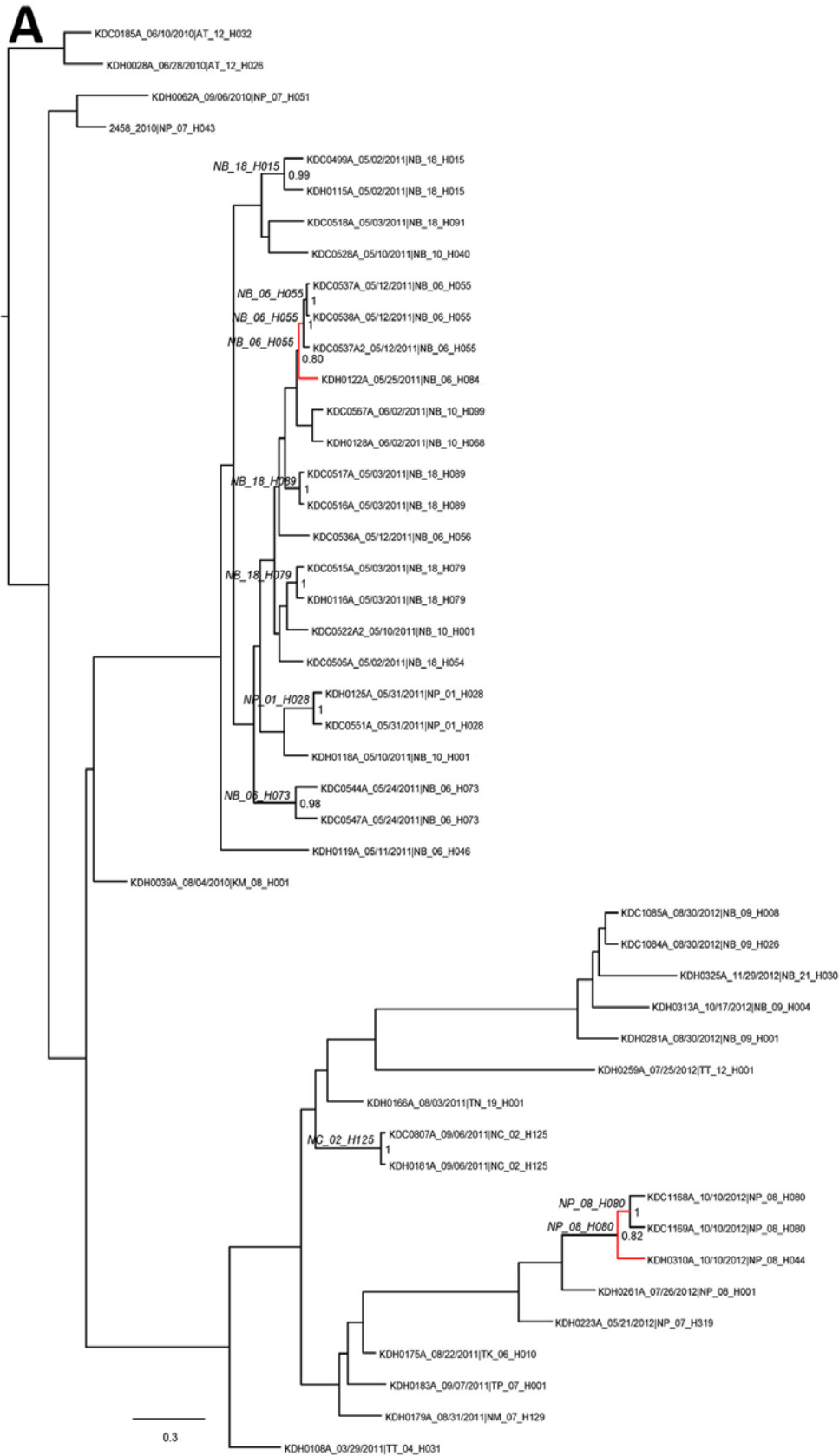
Appendix Figure 1. The number of PCR confirmed dengue cases by year and serotype collected between 2009 and 2012. 2009 samples were not collected across the full year.

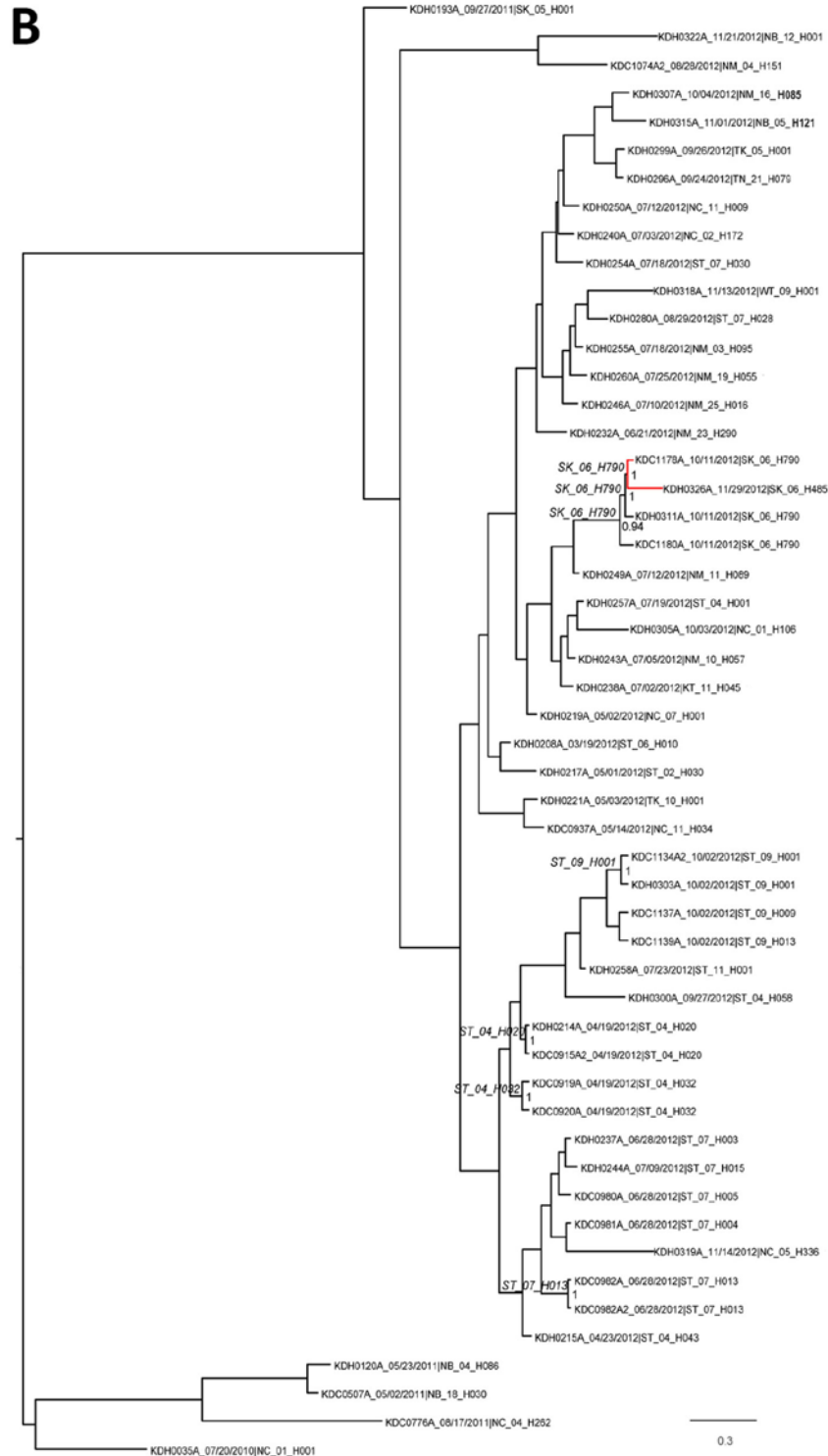
A



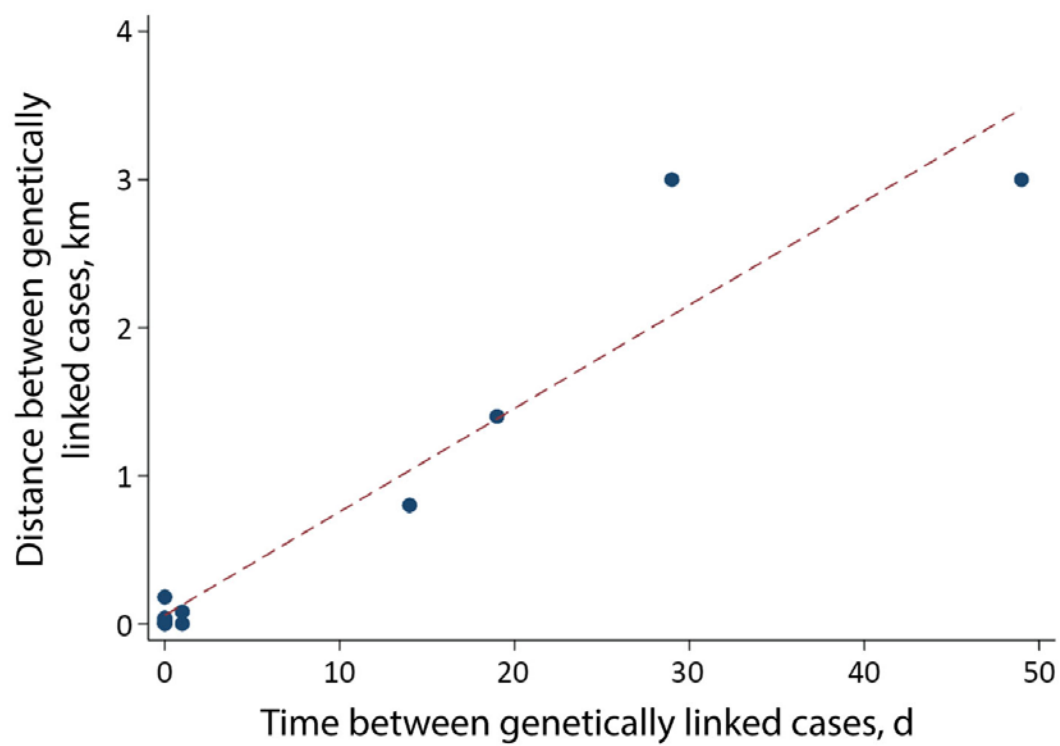


Appendix Figure 2. Maximum clade credibility trees of DENV1 A) Sublineage 1 and B) Sublineage 7, with households as discrete traits. BEAST confirmed transmission clusters/chains are marked red in the tree topology and minor variant confirmed ones are marked green. Purple branch indicates virus from an individual confirmed by both BEAST and minor variant analyses. The three transmission clusters detected by minor variant analyses are marked with green rectangles. All nucleotide positions containing minor variants and the variant composition of those positions are marked next to the respective taxon, shared minor variants are sharing the same color across taxons.





Appendix Figure 3. Maximum clade credibility trees of DENV2 A) Sublineage 2 and B) Sublineage 6, with households as discrete traits. BEAST confirmed transmission clusters/chains are marked red in the tree topology. 95% HPD geo location values >0.8 are shown next to the respective node, and the originating households are marked in italic font.



Appendix Figure 4. Linear regression of transmission-pair distances over sampling time.